

2 Marks

1. Difference between Training data and Testing Data

Training Data	Testing Data
The machine-learning model is trained using training data.	Testing data is used to evaluate the model's performance.
To stop overfitting, training data is utilized.	By making predictions on the testing data and comparing them to the actual labels, the performance of the model is assessed.
Typically larger	Typically smaller
By using the training data, the model can gain knowledge and become more accurate in its predictions.	Until evaluation, the testing data is not exposed to the model.

2. What are the issues in machine learning?

- Inadequate Training Data
- Poor quality of data
- Non-representative training data
- Overfitting and Underfitting
- Getting bad recommendations
- Lack of skilled resources

3. Discuss with examples why machine learning is important.

- Automates tasks: Machine learning helps systems perform tasks without human effort, like detecting fraud in bank transactions.
- Better decisions: It helps businesses make smart decisions, such as recommending products based on customer preferences on platforms like Amazon.
- Handles big data: Machine learning can process huge amounts of data quickly, like Google's search engine, which uses it to find relevant information across the web.

4. What is semi - supervised learning

- Semi-supervised learning is a machine learning technique that uses both labeled and unlabeled data to train AI models.
- It's a combination of supervised and unsupervised learning, and is useful when it's difficult or expensive to get enough labeled data.

Example

- Identifying fraud
 - Semi-supervised learning can be used to train systems that identify cases of fraud or extortion.
- Classifying web content
 - Semi-supervised learning can be used to classify large amounts of web content.

5. Discuss Overfilling in Machine Learning

- In machine learning, overfitting is when a model learns too closely to training data, which can lead to poor performance on new data.
- This is because the model learns the training data's noise and outliers, which reduces its ability to generalise to new data.

Here are some signs of overfitting:

- The model performs well on training data but poorly on new data
- There's a gap between the model's performance on training data and validation data
- The model's learning curve has lower error values

6. Categorize Training set and Test set

The training set trains the machine learning model, allowing it to learn the patterns and relationships within the data.

The test set used after the model has been trained and validated, to provide an unbiased evaluation of the model performance on completely new, unseen data.

A common split for the training and test sets is 70/30 or 80/20.

7. What is classifier in machine learning

- In machine learning, a classifier is an algorithm that automatically assigns data points to a range of categories or classes. Within the classifier category, there are two principal models:
- **Supervised:** In the supervised model, classifiers train to make distinctions between labeled and unlabeled data. This training allows them to recognize patterns and ultimately operate autonomously without using labels.
- **Unsupervised:** Unsupervised algorithms use pattern recognition to classify unlabeled datasets, progressively becoming more accurate.

8. b) Explain both phenomena, with visual representations (learning curves) and techniques like regularization, cross validation to combat them along with its terminologies.

<https://www.javatpoint.com/overfitting-and-underfitting-in-machine-learning>

9. a) Implement the logical operations (AND, OR, NOT) the neuron could model, focusing on its binary decision making using McCulloch-Pitts model.

b) Let us consider an example where the five week's sales data is given as shown below. Apply linear regression technique to predict the 7th and 9th month sales.

X1 (Week)	Y1 (Sales)
1	1.2
2	1.8
3	2.6
4	3.2
5	3.8

<https://www.youtube.com/watch?v=QcPycBZomac>